

CONTENTS

1. Aim of X2DBI	2
2. Interface overview.....	3
3. Database selection.....	4
4. Authentication.....	4
5. Attributes selection	4
6. Filters selection	6
7. Query.....	8
a) Count	8
b) Results.....	8
c) Data export	9
8. Import/Export query.....	10
9. Predefined queries.....	10
10. XML query format	10
11. Interface to browse the database	13

1. Aim of X2DBI

X2DBI is a web application designed as a query-oriented data management system for the construction of complex queries on a database, without knowing its structure.

X2DBI stands for XML to DBI and DBI for DataBase Interface.

Because it is impossible to foresee all possible types of queries in advance, X2DBI can help DB application developers by allowing users to build custom queries on databases themselves. X2DBI does not require any knowledge from users about tables, fields, views and other database issues, and avoids complicated SQL statements. Just include end-user-oriented query builder (X2DBI GUI) into your web-site.

The principle of X2DBI is to convert a database (relational) into **Datasets** which can be accessed *i)* via its standardized web browser interface, *ii)* via web service APIs. Query results are displayed in tabular format and are exportable in a format such as Excel (CSV). This tool needs to be configured by somebody with skills in Perl and SQL scripting. (see <http://sites.google.com/site/x2dbiorg> for installation instructions).

Currently supported RDBMS platforms are MySQL and PostgreSQL.

Funding: Initially developed by the **CBiB**⁽¹⁾ (Bordeaux Bioinformatics Center) and funded within the **ANR PROTICworkshop**⁽²⁾ Project (2008-2010).

Demo Site: <http://services.cbib.u-bordeaux2.fr/x2dbi/>

Availability: <http://sites.google.com/site/x2dbiorg>

Optimized for Mozilla Firefox, Google Chrome & Internet Explorer 9

Licence: X2DBI software is completely Open Source, licensed under the GNU General Public License, and freely available to anyone to use and redistribute without restriction.

Contacts:

- Daniel Jacob: daniel.jacob@u-bordeaux2.fr
- Laurent Gil: lgil@ebi.ac.uk

(1) <http://www.cbib.u-bordeaux2.fr/en>

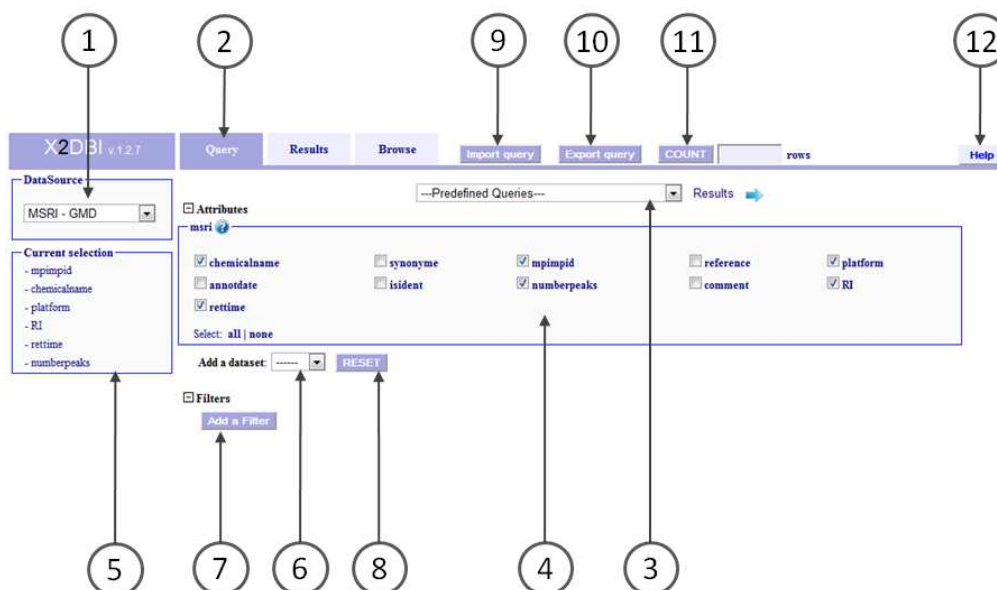
(2) <http://moulon.inra.fr/index.php/en/traverse-team/atelier-de-bioinformatique/projects/97>

2. Interface overview

You will find below a short description of the web interface of X2DBI. The web address to access to your X2DBI is as follows:

http://<your_web_site>/x2dbi/

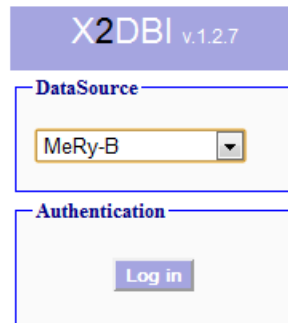
For example: <http://www.cbib.u-bordeaux2.fr/x2dbi/>



N°	Description
1	Datasource selection list. (Databases)
2	Tabs for Query building (Attributes and Filters), for query launching, for results consulting (Results) and for browsing the database (Browse).
3	Selection list of predefined queries
4	Selection of attributes to include in the query (result columns)
5	List of selected attributes, displayed in the same order than within the result table.
6	Datasets selection list for extending the general query.
7	filters to add constraints on the query (query conditions)
8	Reset button for all the datasets and attributes.
9	Button to import a query from an XML file (previously exported or manually generated)
10	Button to export the query built (as an XML file).
11	Calculate the number of results returned for the query. The number of results is displayed in the textbox beside.
12	HELP button

3. Database selection

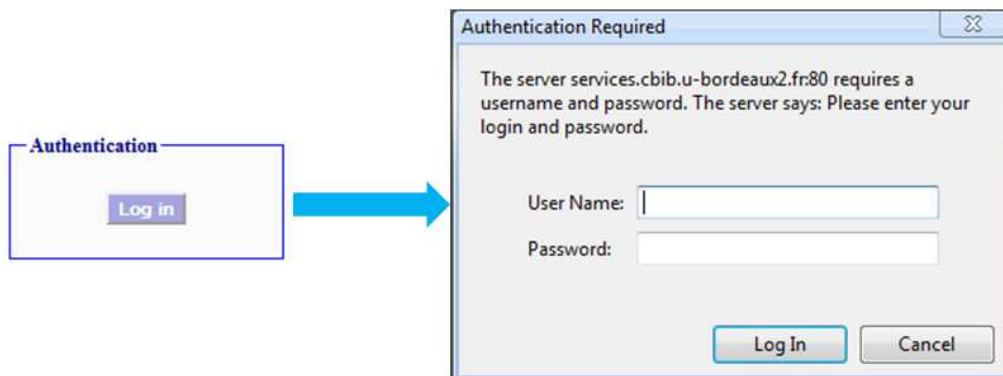
X2DBI can provide access to several databases (but only one at a time). So, the first step is to select the database you want to query.



The image shows the X2DBI v1.2.7 interface. It has a purple header bar with the text 'X2DBI v1.2.7'. Below the header, there are two sections. The first section is labeled 'DataSource' and contains a dropdown menu with 'MeRy-B' selected. The second section is labeled 'Authentication' and contains a 'Log in' button.

4. Authentication

The private databases are governed by access rights. To can access to the private data of a database, you must enter a username and password (corresponding to a database account), so that X2DBI can authenticate you to the chosen database.



The diagram illustrates the authentication process. On the left, there is a box labeled 'Authentication' containing a 'Log in' button. A blue arrow points from this button to a dialog box titled 'Authentication Required'. The dialog box contains the following text: 'The server services.cbib.u-bordeaux2.fr:80 requires a username and password. The server says: Please enter your login and password.' Below this text are two input fields: 'User Name:' and 'Password:'. At the bottom of the dialog box are two buttons: 'Log In' and 'Cancel'.

By default, X2DBI provides you access to the public data of the selected database. So, you don't need to use the authentication form if you only want to access to the public data.

5. Attributes selection

Then, you can select the attributes, from the different datasets, on to the **Attributes** panel in the **Query** tab. They will be displayed in the results table, as columns. To ease the selection of

attributes, they are grouped in logic datasets. Usually, they are gathered around a central entity and following a logical process (e.g. OMICs experiments). A dataset is always displayed by default. Generally, it concerns the attributes describing the main entities (like experimental design).

☐ Attributes

samples ?

<input type="checkbox"/> taxon id	<input type="checkbox"/> species	<input type="checkbox"/> PS accession	<input type="checkbox"/> tissue	<input type="checkbox"/> PG accession
<input type="checkbox"/> developmental stage	<input type="checkbox"/> age	<input type="checkbox"/> env condition	<input type="checkbox"/> genotype	<input type="checkbox"/> background
<input type="checkbox"/> study type	<input type="checkbox"/> project id	<input type="checkbox"/> project name	<input type="checkbox"/> experiment id	<input type="checkbox"/> experiment name
<input type="checkbox"/> individual	<input type="checkbox"/> sample name			

Select: all | none

You can add other datasets using the selection box (**Add a dataset**), located below the last dataset field.

By clicking "**all**" (at the bottom left of each dataset), all attributes of the group will be checked. Conversely, by clicking "**none**", all attributes of the group will be unchecked.

samples ?

<input type="checkbox"/> taxon id	<input type="checkbox"/> species	<input type="checkbox"/> PS accession	<input type="checkbox"/> tissue	<input type="checkbox"/> PG accession
<input type="checkbox"/> developmental stage	<input type="checkbox"/> age	<input type="checkbox"/> env condition	<input type="checkbox"/> genotype	<input type="checkbox"/> background
<input type="checkbox"/> study type	<input type="checkbox"/> project id	<input type="checkbox"/> project name	<input type="checkbox"/> experiment id	<input type="checkbox"/> experiment name
<input type="checkbox"/> individual	<input type="checkbox"/> sample name			

Select: all | none

compounds ?

<input type="checkbox"/> compound name	<input type="checkbox"/> accession id	<input type="checkbox"/> meryb id	<input type="checkbox"/> user synonym	<input type="checkbox"/> compound description
<input type="checkbox"/> nmr shift	<input type="checkbox"/> nmr shift description	<input type="checkbox"/> nmr multiplicity	<input type="checkbox"/> experiments set name	

Select: all | none

Add a dataset:

However, you cannot display any group of attributes at any place: they are related to each other with an order defined by the website administrator who setup the application.

X2DBI v.1.2.7

DataSource

Authentication

Current selection

- meryb id
- compound name
- user synonym
- nmr multiplicity
- nmr shift
- nmr shift description
- compound description

The list of selected attributes is dynamically displayed in the left panel (field "Current selection"), below the authentication field. It shows you what will be order of the columns in the result table.

Thus, the first column of the result table will be "meryb_id", the second column will be "compound name", and so on.

Help - subset "samples"

Attribute	Description
taxon id	NCBI Taxon ID (TO)
species	NCBI Taxon Name (TO)
PS accession	Tissue Accession (POC)
tissue	Tissue Name (POC)
PG accession	Growth and Dev. Stage Accession (POC)
developmental stage	Growth and Dev. Stage (POC)
age	Age (dpa)
env condition	Treatment
genotype	Genotype (CV)
background	Genotype Background (CV)
study type	Condition of Study (EO)
project id	Project ID
project name	Project Name
experiment id	Experiment ID
experiment name	Experiment Name
individual	Individual Name
sample name	(Plant) Sample Code or Name

Example of help form (?), appearing as a popup:

6. Filters selection

Once your attributes are selected, you can use filters to add constraints on your query by clicking on the “Add a Filter” button on to the **Filters** panel in the **Query** tab. This step is optional.

If you want to limit your results to a certain range/values of attributes, you can create one or more filters.

By default, a blanked form appears:

Filters

Filter 1

----- -- exact match

Add a Filter

It is framed in red to indicate that in this state, the constraint will not be included in your query. Then, you need to fill the form (the selection field on the left and the text field on the right), if you want your constraint to be taken into account. It switches in blue to indicate that the constraint will be integrated into your query:

Filters

Filter 1

project name -- exact match Tomato- Metabolomics - 2006 Delete filter

For some attributes, an auto completion is enabled in the text field if the following conditions are true:

- The values of the attribute can be browsed in the *Browse* tab.
- The *exact value* is selected in the select type field.

You can add another filter by clicking the “**Add a Filter**” button.

You also can delete a filter by using the “**Delete**” button, in each form. This button appears only if at least 2 filter forms are displayed.

The image shows two filter forms, Filter 1 and Filter 2. Each form contains a dropdown menu for the attribute, a separator '--', a dropdown for the match type (set to 'exact match'), a text input field for the value, and a 'Delete filter' button.

Filter 1: attribute 'project name', value 'Tomato- Metabolomics - 2006'.

Filter 2: attribute 'compound name', value 'Chlorogenate'.

Once an attribute set is selected in the **Query** tab, you can filter the query on any attribute of this set, even if this attribute is not selected in the query.

There are several types of filters:

Type	Description
exact match	Filter to an exact numeric or alphanumeric value.
between	Filter between two numeric values.
like	Filter on a numeric or alphanumeric value (e.g. XX) which begins by XX, contains XX or ends by XX. For an advanced use, you can write the value directly with the character '%': <ul style="list-style-type: none"> • %XX : search for values ending with XX. • XX%: search for values beginning with XX. • %XX%: search for values containing XX.
in set - numeric	Filter on a set of numeric values separated by a ";". It also filters ranges of values with the separator ":" Ex: "1.1:1.2;3.5:3.6" defines the two ranges ([1.1 ... 1.2] and [3.5 ... 3.6])
in set - character	Filter on a set of string values separated by a ";". For an advanced use, you can write strings with the character '%', like the "like" filter: <ul style="list-style-type: none"> • %XX : search for values ending with XX. • XX%: search for values beginning with XX. • %XX%: search for values containing XX. Ex: "%ose;%ase" defines two groups, respectively strings ending by "ose" and "ase"
greater than	Filter on values strictly greater (>) than the given value.
less than	Filter on values strictly less (<) than the given value.

NOT	Negation which inverts the filter conditions
------------	--

7. Query

a) Count

Before submit the query and display all the results you can, as a first step, see the number of results found by clicking on the button *COUNT*. The number of results can be seen in the upper right of the page:



If you change the attributes or the filters, after submitting a *COUNT*, the number into the "count field" will be coloured in red:



It indicates that the count result does not correspond necessarily to the current query and need to be updated. To obtain the number of results returned by your new query, simply click again to the **COUNT** button.

Note: You can stop a **COUNT** query by clicking the *STOP* button that appears at the top right of the page.

b) Results

Once attributes were selected and filters were added, you can submit the application by clicking on the *Launch query* button (in the top of the Results tab).



You can sort your results table, by selecting the column to sort through the field "Sortby".

Note: In the command line version of X2DBI, it is possible to sort the table on several columns, inserting multiple *OrderBy* tags in the XML query file (see [section 9](#)).

To launch the query, simply click on *Launch query*, just below the tabs bar.

After a few time, results appear as a table with, in columns, the attributes selected in the

Query tab:

Exports : CSV													
Rows : 4 found													
taxon id	species	project id	project name	experiment id	experiment name	ps accession	tissue	pg accession	developmental stage	age	genotype	background	study type
4081	Lycopersicon esculentum	10	Greenhouse Tomato - Hortis - Metabolomics 2009	17	Palmiro without nutrient solution recycling	PO:0009084	pericarp	PO:0007038	FR.04 fruit ripening complete		Palmiro	WT	green house study
4081	Lycopersicon esculentum	10	Greenhouse Tomato - Hortis - Metabolomics 2009	18	Palmiro and Clotilde	PO:0009084	pericarp	PO:0007038	FR.04 fruit ripening complete		Clotilde	WT	green house study
4081	Lycopersicon esculentum	10	Greenhouse Tomato - Hortis - Metabolomics 2009	18	Palmiro and Clotilde	PO:0009084	pericarp	PO:0007038	FR.04 fruit ripening complete		Palmiro	WT	green house study
4081	Lycopersicon esculentum	10	Greenhouse Tomato - Hortis - Metabolomics 2009	19	Palmiro with nutrient solution recycling	PO:0009084	pericarp	PO:0007038	FR.04 fruit ripening complete		Palmiro	WT	green house study

Note: You can stop a query by clicking the *STOP* button that appears at the top right of the page.

If an URL link has been associated with some attributes, values of these attributes appear as such, and you can click on them to access the linked resources.

c) Data export

It is possible to export results in CSV format (button on the horizontal bar above the results table), for processing with other tools:

```
#4 rows
taxon_id;species;project_id;project_name;experiment_id;experiment_name;ps_accession;tissue;pg_accession;developmental_stage;age;genotype;background;study_type
@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/TAXO/@4081;Lycopersicon esculentum;@http://www.cbib.u-bordeaux2.fr/MERYB/res/project/@10;Greenhouse Tomato - Hortis - Metabolomics 2009;@http://www.cbib.u-bordeaux2.fr/MERYB/res/experiment/@17;Palmiro without nutrient solution recycling;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PS/@PO:0009084;pericarp;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PG/@PO:0007038;FR.04 fruit ripening complete;;Palmiro;WT;green house study
@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/TAXO/@4081;Lycopersicon esculentum;@http://www.cbib.u-bordeaux2.fr/MERYB/res/project/@10;Greenhouse Tomato - Hortis - Metabolomics 2009;@http://www.cbib.u-bordeaux2.fr/MERYB/res/experiment/@18;Palmiro and Clotilde;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PS/@PO:0009084;pericarp;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PG/@PO:0007038;FR.04 fruit ripening complete;;Clotilde;WT;green house study
@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/TAXO/@4081;Lycopersicon esculentum;@http://www.cbib.u-bordeaux2.fr/MERYB/res/project/@10;Greenhouse Tomato - Hortis - Metabolomics 2009;@http://www.cbib.u-bordeaux2.fr/MERYB/res/experiment/@19;Palmiro with nutrient solution recycling;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PS/@PO:0009084;pericarp;@http://www.cbib.u-bordeaux2.fr/MERYB/res/ontology/PG/@PO:0007038;FR.04 fruit ripening complete;;Palmiro;WT;green house study
```

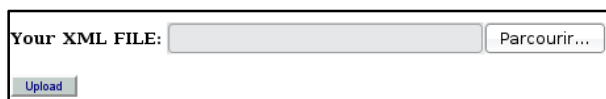
You can then return to the Query tab to modify and/or refine your search. To launch your new query, simply click again on the **Launch** button in the **Results** tab.

Note: In the case of URLs links have been associated with some attributes; when the query is submitted through the command line interface (CLI), the first part of the URL appears between two '@' characters as shown here. Otherwise, when exporting through the Web interface (GUI), this part of URL will not be included.

8. Import/Export query

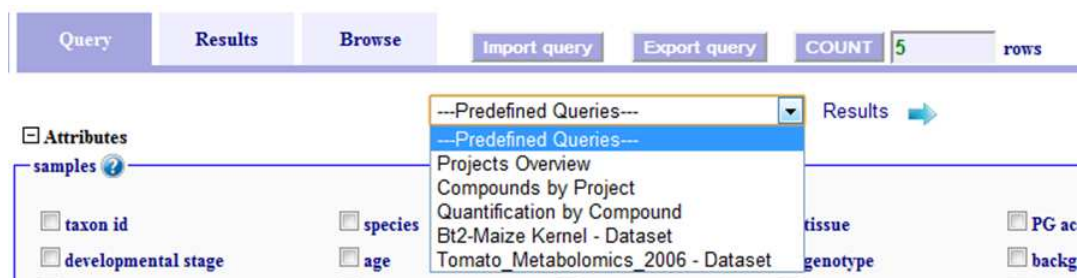
In X2DBI, you can save your query (filters + attributes) in XML format using the **“Export query”** button.

Then, you can load your XML query in X2DBI, using the **“Import query”** button.



9. Predefined queries

If you have any queries frequently used, it is best to ask the website administrator who setup the application to put it in the list of predefined queries. One can see the predefined queries as shortcuts for importation which internally works in the same way.



Thus, you can simply by selecting a query in this list, set the corresponding datasets and attributes within the interface, possibly make some modification (e.g. values of filter conditions) and submit it.

10. XML query format

This section concerns a technical part of the software. The aim of this section is to describe the structure and the content of the XML query file, if you want to create or edit it yourself before importing directly the query in X2DBI.

A XML query has the following structure:

Query → Dataset → Attribute, Filter

Query → OrderBy

The Filter tag can contain different XML attributes, depending on the type of filter chosen by the user:

Filter → min,max (for the filter "*between*")

Filter → item (pour les for the filters "*in set - numeric*" and "*in set - character*")

Here is a XML query template

```
<Query virtualSchemaName="default" datasetConfigVersion="1.2.6">
  <Dataset name="..." interface="default">
    <Attribute name="..." />

    <Attribute name="..." />

    ...

    <Filter type="..." name="..." value="..." />
    ...
  </Dataset>
  <Dataset name="..." interface="default">
    ....
  </Dataset>
  <OrderBy name="..." />
</Query>
```

The **Query** tag defines the query itself. The **datasetConfigVersion** setting of 1.2.6 indicates that you are using XML compatible with the dataset configuration being queried. There can be one or multiple **dataset** tags, depending on the query. There can be multiple **Attribute** and **Filter** tags, each indicating the name of an attribute or filter to apply. Filter tags also accept a value to apply to the filter. Attributes in the result columns returned will be sorted first by the order of the Dataset tags in which they appear, then by the order of the Attribute tags within the Dataset tags.

Each entity has XML attributes:

Tag	List of XML attributes	Number
Query (#)	- <u>virtualSchemaName</u> - <u>datasetConfigVersion</u>	1
Dataset (#)	- <u>name</u> : subset name - <u>interface</u> : has the value "default"	1 to n
Attribute (#)	- <u>name</u> : name of the selected attributes. Unlike the interface, attribute names do not have space. They are replaced by the character "_". Example: "project name" in the web interface become "project_name" in the XML file.	1 to n
Filter (#)	- <u>type</u> : filter category. The categories are: - <i>attribute</i> as "exact match" - <i>like</i>	0 to n

	<ul style="list-style-type: none"> - <i>interval</i> as "between" - <i>set</i> as "in set – numeric" - <i>setlike</i> as "in set – character" - <i>greater</i> as "greater than" - <i>less</i> as "less than" <ul style="list-style-type: none"> - <u>name</u>: attribute name - <u>value</u>: filter value (if the filter is different than <i>interval</i>, <i>set</i> and <i>setlike</i>) 	
OrderBy (#)	<ul style="list-style-type: none"> - <u>name</u>: attribute name to sort <p>*only the use of the command line for the XML query can take into account more than one tag OrderBy.</p>	0 to n*
min	<ul style="list-style-type: none"> - <u>value</u>: filter minimum value * if the filter type is <i>interval</i> 	1*
max	<ul style="list-style-type: none"> - <u>value</u>: filter maximum value * if the filter type is <i>interval</i> 	1*
item	<ul style="list-style-type: none"> - <u>value</u>: filter value * if the filter type is <i>setlike</i> 	1 to n*

See below an example to illustrate the file structure of an XML query:

```
<Query virtualSchemaName="default" datasetConfigVersion="1.2.6">
  <Dataset name="samples" interface="default">
    <Filter type="attribut" name="project_name" value="Tomato- Metabolomics - 2006"/>
  </Dataset>
  <Dataset name="compounds" interface="default">
    <Attribute name="meryb_id"/>
    <Attribute name="compound_name"/>
    <Attribute name="user_synonym"/>
    <Attribute name="nmr_multiplicity"/>
    <Attribute name="nmr_shift"/>
    <Attribute name="nmr_shift_description"/>
    <Attribute name="compound_description"/>
  </Dataset>
  <OrderBy name="compound_name"/>
</Query>
```

You can see 2 datasets (**Dataset**), each containing one or more attributes (**Attribute**). There is only one constraint (**Filter**), placed on the first dataset (**Dataset name="samples"**). The data sorting is done according to the species column (**OrderBy name="species"**).

We have to read the XML query exemple above as:

```
SELECT (Attributes)
  meryb_id, compound_name, user_synonym, nmr_multiplicity,
  nmr_shift, nmr_description, compound_description
WHERE (Filters)
  projet_name = "Tomato - Metabolomics - 2006"
ORDER BY
  compound_name
```

It gives the results as below:

Exports : **CSV**

Rows : 43 found

meryb id	compound name	user synonym	nmr multiplicity	nmr shift	nmr shift description	compound description
MRB7	4-Aminobutanoate	GABA	Multiplet	1.92	C3H2	
MRB7	4-Aminobutanoate	GABA	Triplet	2.3	C2H2	
MRB7	4-Aminobutanoate	GABA	Triplet	3.01	C4H2	
MRB21	5-Oxoproline	pyroglutamic acid	Doublet of doublets	4.18	C2H	
MRB5	Acetylcholine	Acetylcholine	Singlet	3.22	N-(C(5)H3)3	
MRB63	Alanine	Alanine	Doublet	1.48	C(3)H3	
MRB85	Asparagine	Asparagine	Multiplet	2.92	C(3)H2	
MRB86	Aspartate	Aspartate	Multiplet	2.81	1/2(C(3)H2)	

cutted...

11. Interface to browse the database

X2DBI can also browse the database, by subset (buttons on the horizontal bar below the tabs) and by attribute in order to display and fetch the list of values of an attribute in the database:

Query Results Browse Import query Export query COUNT rows

Query Results

Subsets : compounds gcmsbucket nmrbucket nmrpeaklist quantifications samples

Subset > samples

Links: Subset(s) NMRBucket Compounds

Attributes:

- taxon id
- species
- PS accession
- tissue
- PG accession
- developmental stage
- age
- env condition
- genotype
- background
- study type
- project id
- project name
- experiment id
- experiment name
- individual
- sample name

Attribute > species

Subset	samples
Description	NCBI Taxon Name (TO)
Values	Arabidopsis thaliana Beta vulgaris Brassica oleracea var. italica Cucumis melo Daphne gnidium Elaeis guineensis Fragaria x ananassa Lycopersicon esculentum Medicago truncatula Oryza sativa Ostreococcus tauri Phoenix dactylifera Prunus persica Vanilla planifolia Vitis vinifera Zea mays

To access to a dataset, you can click the group of interest in the bar at the top of the image

above. The row "Links" in the table represents the datasets (e.g. Compounds) related to the displayed dataset (samples).

Each attribute of the list is clickable. For some attributes, the list of values is not displayed (attributes not "Browsable"). This concerns in particular the analyzed data, for the sake of confidentiality of private data but also because of the big amount of data to display.